Interpreting the SF-12

One of the problems commonly reported by those who have used the SF12 to measure health status is that the Physical and Mental Health Composite Scale Scores have little intuitive meaning. For instance, what does a score of 42.5 mean? The following section is intended to give the reader a better understanding of the meaning of different scale values by examining the questionnaire items that are used to generate the composite scale scores, and the range of scale scores in Utah.

Because a person's composite scale score traditionally differs over the life span (with age, scores decrease for physical health and increase for mental health), this section also provides mean scores for ten-year age groups. When interpreting a person's score, one should use their age-specific group mean score as a reference point. Scores higher than the mean indicate that a person has better health status than most other persons his or her age, while scores lower than the age group mean indicate poorer health status than most other persons of the same age.

Finally, this section describes the Age-Specific Mean Difference Score. This difference score is the amount by which a person's score differs from their age group mean score. That is, a person with a difference score of -5.5 will score 5.5 points lower than other persons their age -- an indication of somewhat poorer health. The difference score is perhaps the most intuitive way to judge a persons score. Looking at a difference score, it is immediately clear whether a person is more healthy or less healthy than other persons in his or her comparison group. In addition, difference scores can be compared across age groups. That is, a score of -5.5 means the same thing, regardless of a person's age. Finally, difference scores have an additional advantage, because they can be characterized as "average," "below average," or "above average." A finding that a person is "below average" is immediately interpretable, especially compared to just knowing that his or her score is 42.5. Age-Specific Mean Difference Scores have been computed for both the Physical and Mental Health Composite Scales.

Responses to the 12 Survey Questions

- **The SF-12 consists of 12 survey questions, shown in the following bar charts. The percentages reflect the distribution of responses of the 6,131 survey respondents, weighted to reflect the age, SES, Hispanic status, and geographic distribution of adults age 18 or over in Utah.**

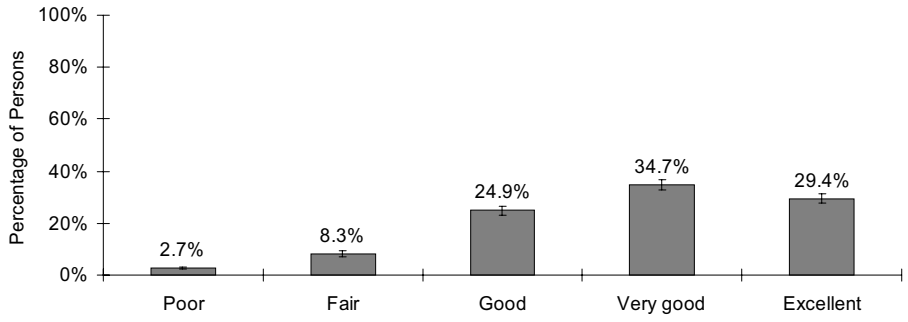
- **The 12 bar charts below have been constructed with the highest (healthiest) scores on the right and the lowest scores on the left.**

- **The 12 items include questions about both physical and mental health. A weighted sum of all 12 items is used in creating an individual's physical and mental health composite scale scores. The difference in the two scales depends on how much weight is given to each item. (For additional information, see Ware, Kosinski & Keller, 1995, and the technical appendix at the back of this report.)**

- **In the pages that follow, the 12 items have been organized according to the eight subdomains of health status measured by the SF-12.**

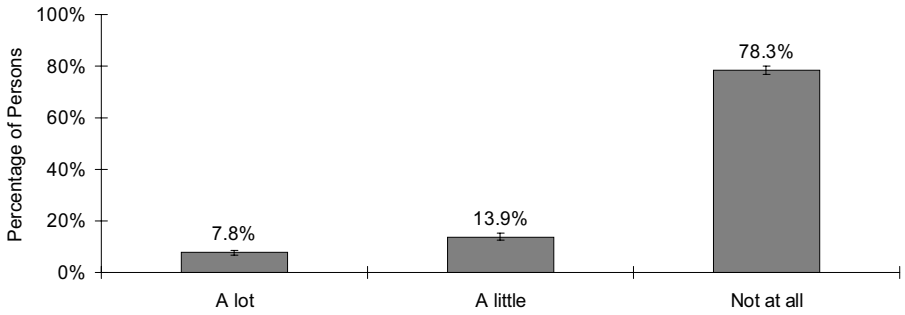
General Health Subdomain

In general, would you say your health is excellent, very good, good, fair, or poor?

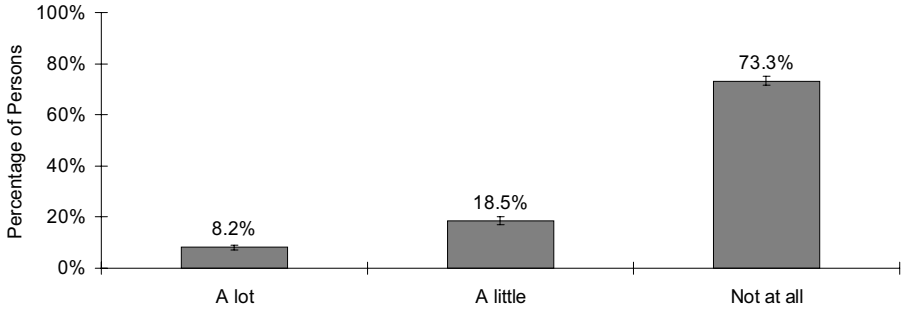


Physical Functioning Subdomain

Does YOUR HEALTH NOW LIMIT YOU IN MODERATE ACTIVITIES, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf? Would you say you are limited a lot, a little, or not at all?

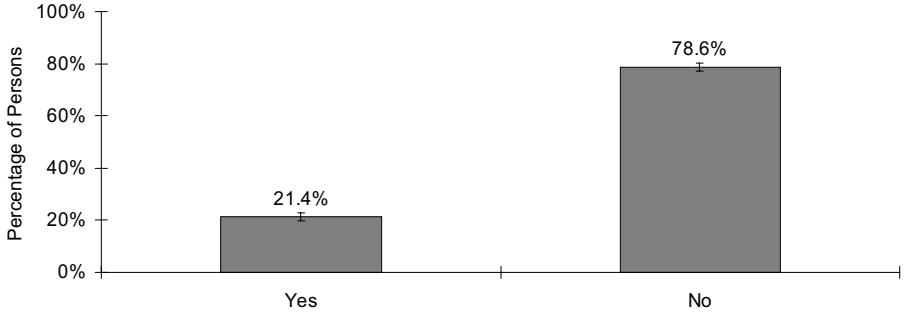


How about CLIMBING SEVERAL FLIGHTS OF STAIRS? Would you say your health limits you a lot, a little, or not at all?

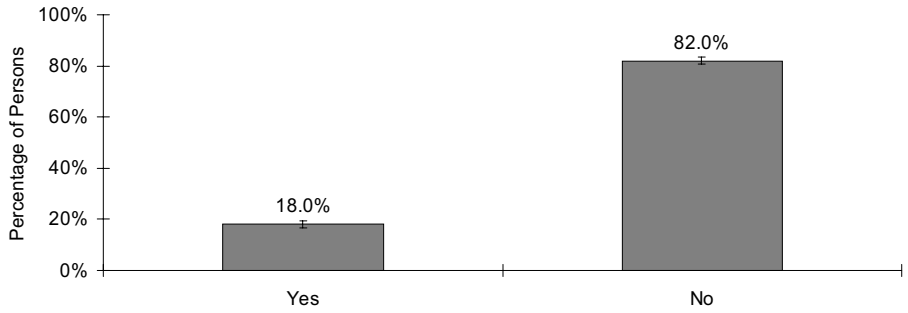


Role Functioning (Physical) Subdomain

Thinking about the past four weeks, have you ACCOMPLISHED LESS than you would like AS A RESULT OF YOUR PHYSICAL HEALTH?

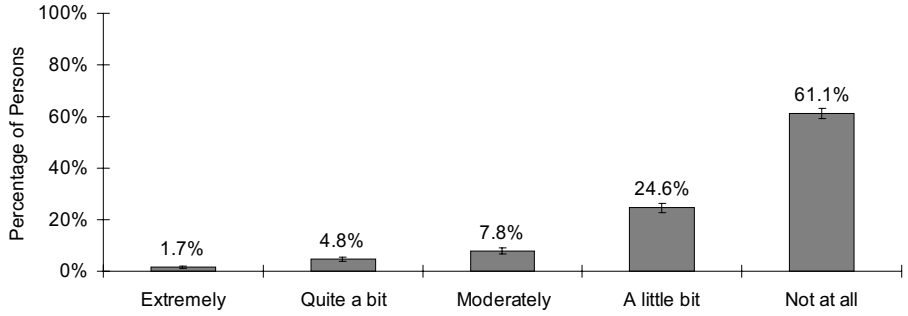


During the past four weeks, were you limited in the KIND of work or other activities you could do as a result of your physical health?



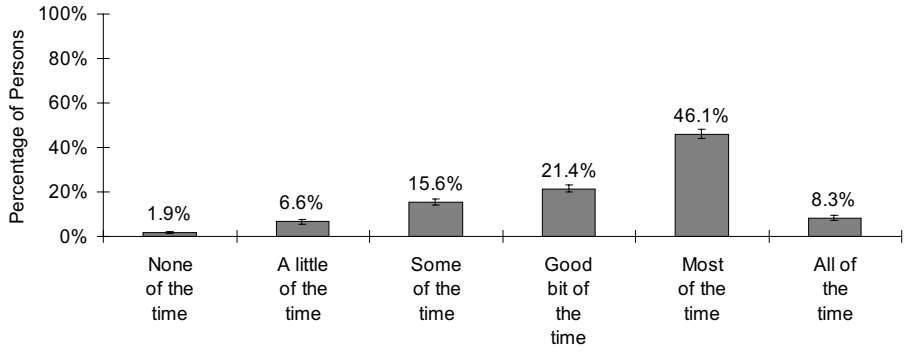
Bodily Pain Subdomain

During the past four weeks, how much did PAIN interfere with your normal work including both work outside the home and housework, would you say *(read responses)*?



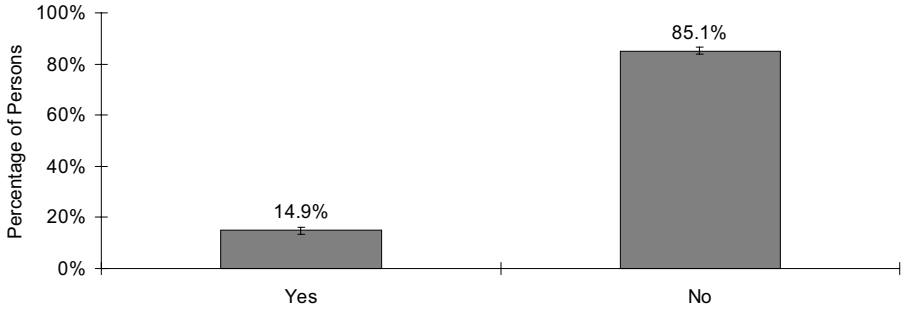
Vitality Subdomain

How much of the time during the PAST FOUR WEEKS did you have a lot of energy? Would you say *(read responses)*?

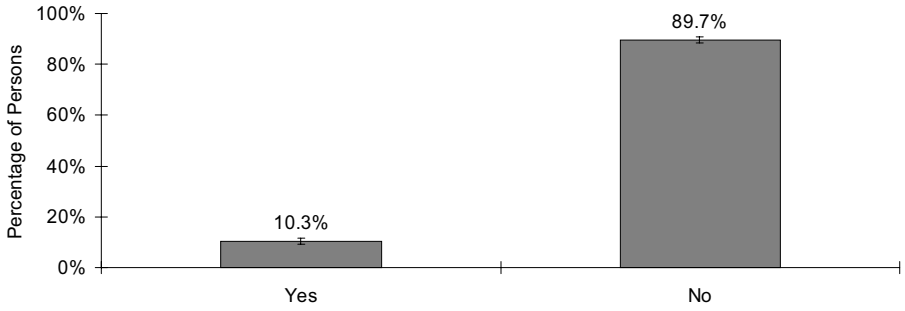


Role Functioning (Emotional) Subdomain

In the past four weeks, did you ACCOMPLISH LESS than you would like AS A RESULT OF AN EMOTIONAL PROBLEM, such as feeling depressed or anxious?

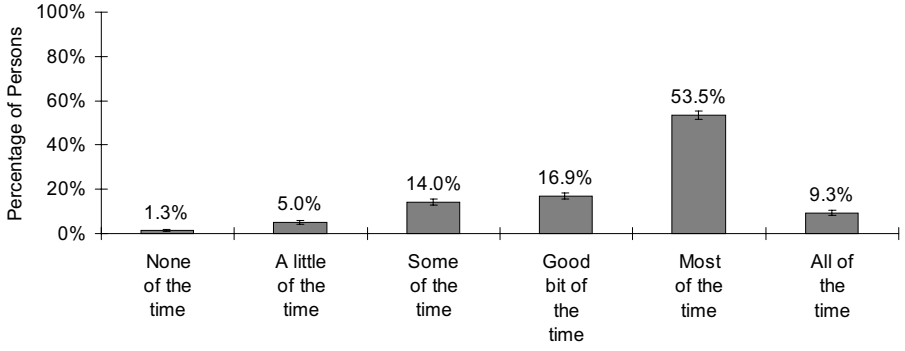


During the last four weeks, did you have trouble doing work or other activities as CAREFULLY as usual AS A RESULT OF AN EMOTIONAL PROBLEM, such as feeling depressed or anxious?

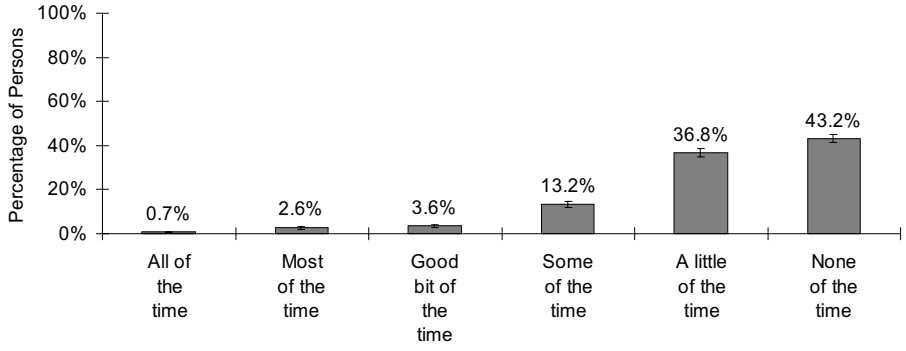


Mental Health Subdomain

How much of the time during the past four weeks have you felt calm and peaceful? Would you say *(read responses)*?

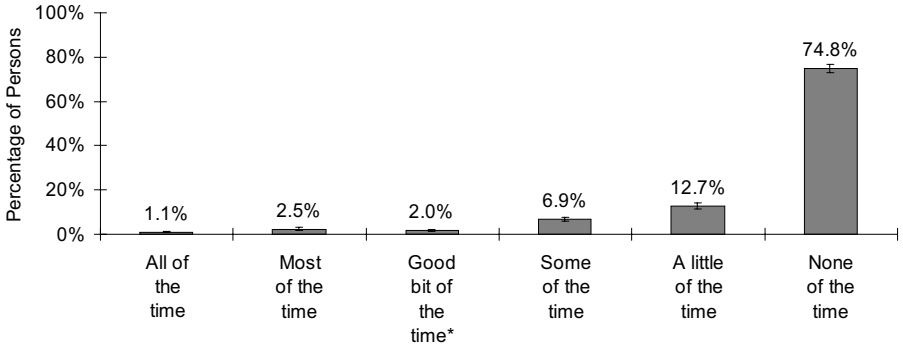


How much of the time during the past four weeks have you felt downhearted and blue? *(If necessary, read responses)*



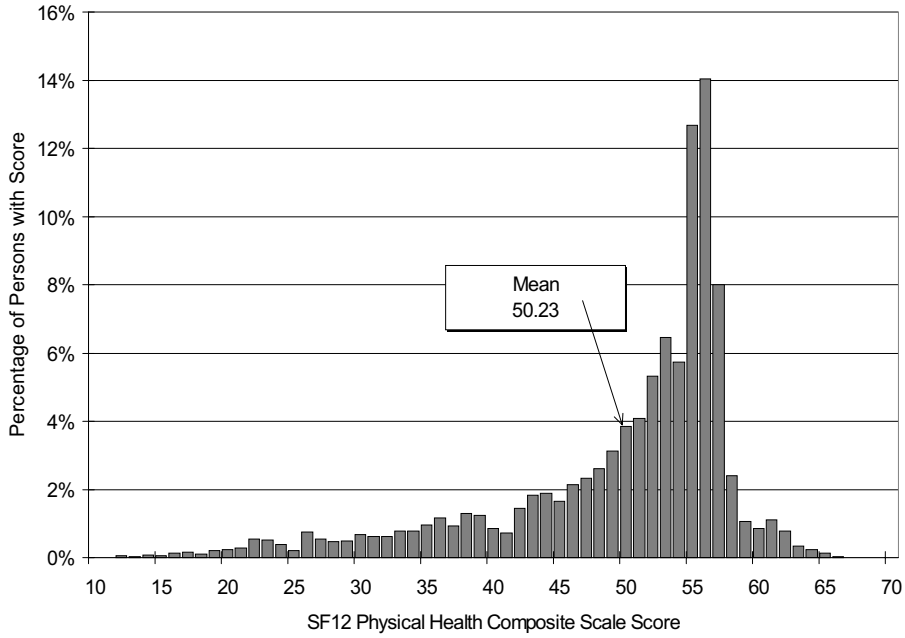
Social Functioning Subdomain

During the last four weeks, how much of the time has your PHYSICAL HEALTH OR EMOTIONAL PROBLEMS interfered with your social activities, like visiting with friends, relatives, etc.? *(If necessary, read responses)*

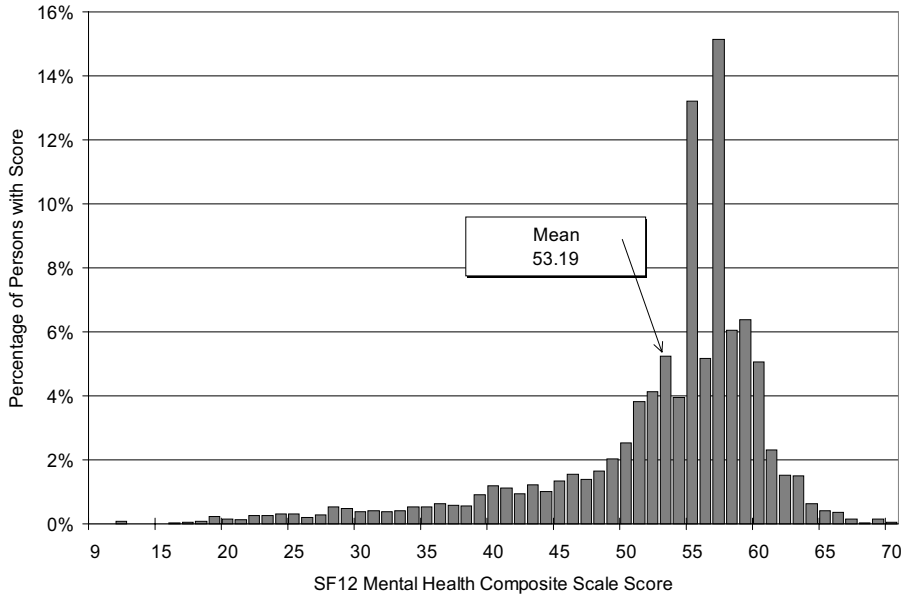


* This response alternative was unintentionally added to the social functioning item. Its inconclusion precludes absolute comparability to other SF-12 results. In practice, however, our analyses and comparisons with other Utah survey results indicate that its inclusion had a negligible effect on the data distribution and on SF-12 summary scale values.

Medical Outcomes Study SF12 Physical Health Composite Scale Scores: Utahns Age 18 or Over, 1996



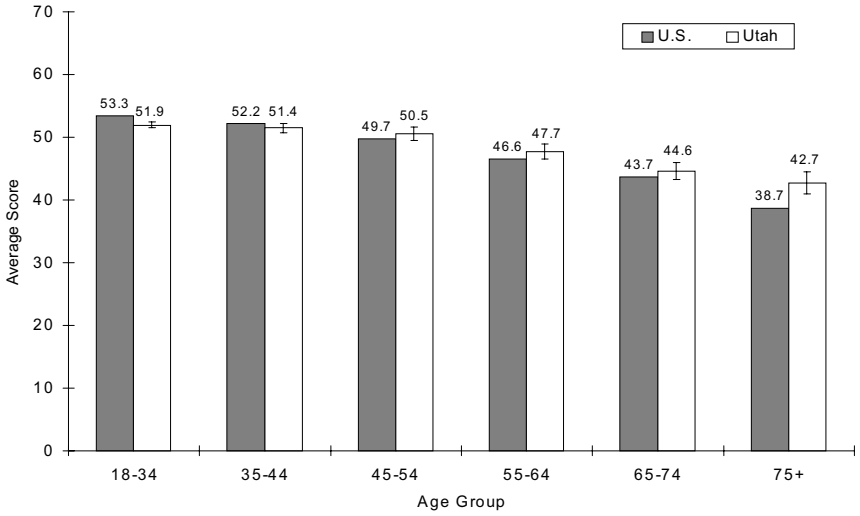
Medical Outcomes Study SF12 Mental Health Composite Scale Scores: Utahns Age 18 or Over, 1996



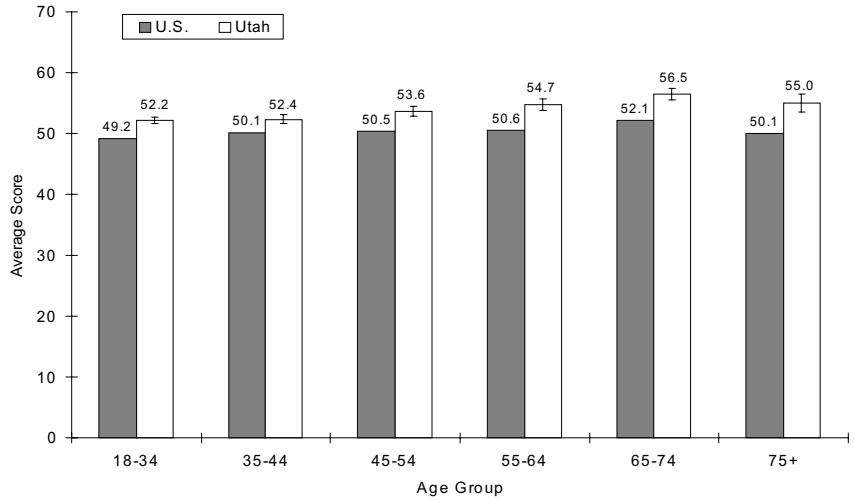
- **Both Physical and Mental Health Composite Scales combine the 12 items in such a way that they compare to a national norm of a mean score of 50.0 and a standard deviation of 10.0.**

- **In Utah, the mean (average) scores are slightly above 50. Utahns scored significantly above the U.S. population on the mental health summary measure, but not on the physical health summary measure.**

Comparison of Utah and U.S. Physical Health Composite Scale (PCS12) Scores: Utah 1996, U.S. 1995*



Comparison of Utah and U.S. Mental Health Composite Scale (PCS12) Scores: Utah 1996, U.S. 1995*



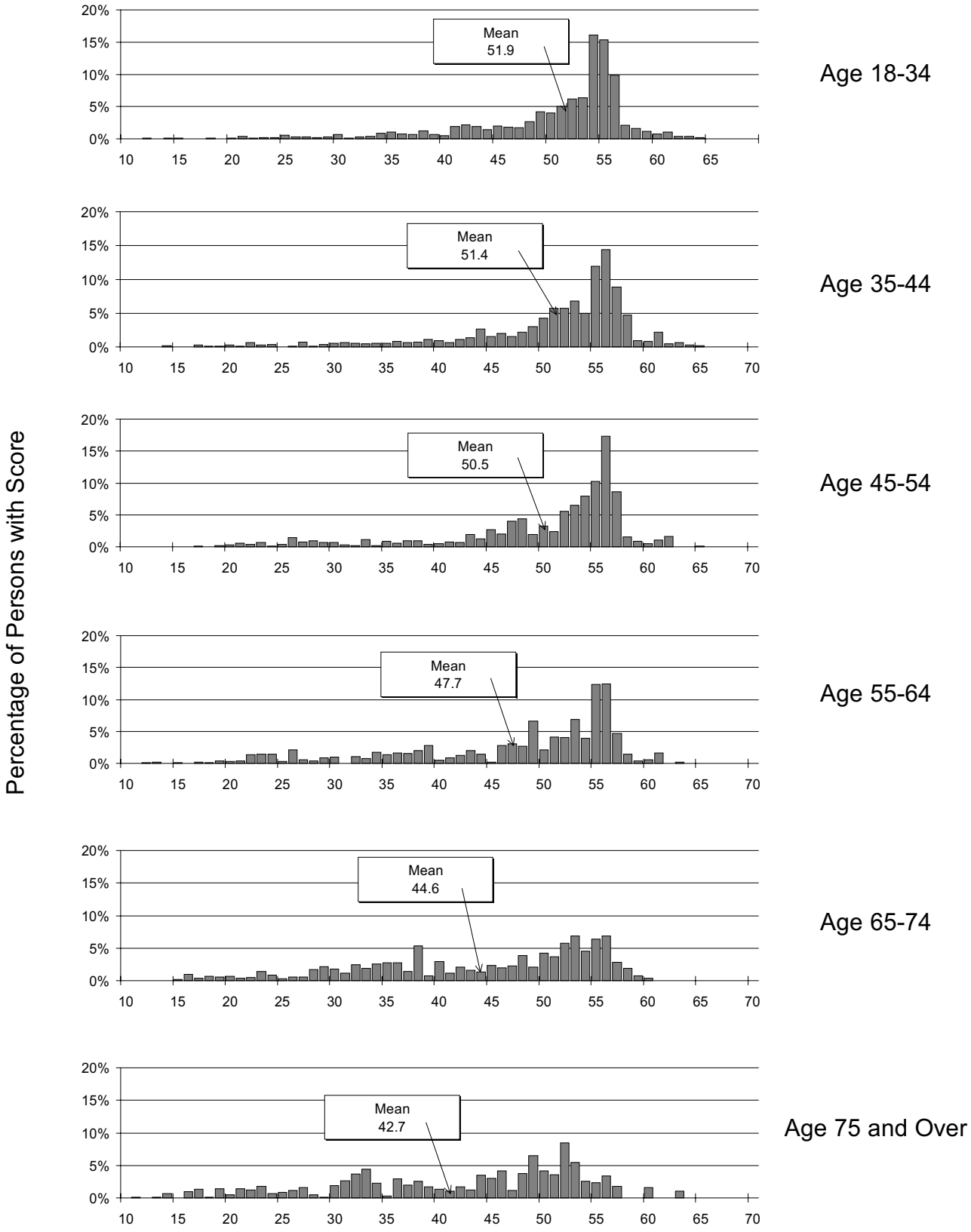
*U.S. norms as reported in Ware, Kosinski, & Keller, 1995.

- **With age, persons tend to score lower on the physical health scale, but better on the mental health scale. This is a robust finding that can be found in other populations, and using other measurement tools. Since there are systematic age differences in scores, it is important to interpret a person's score in the context of other persons near their same age.**

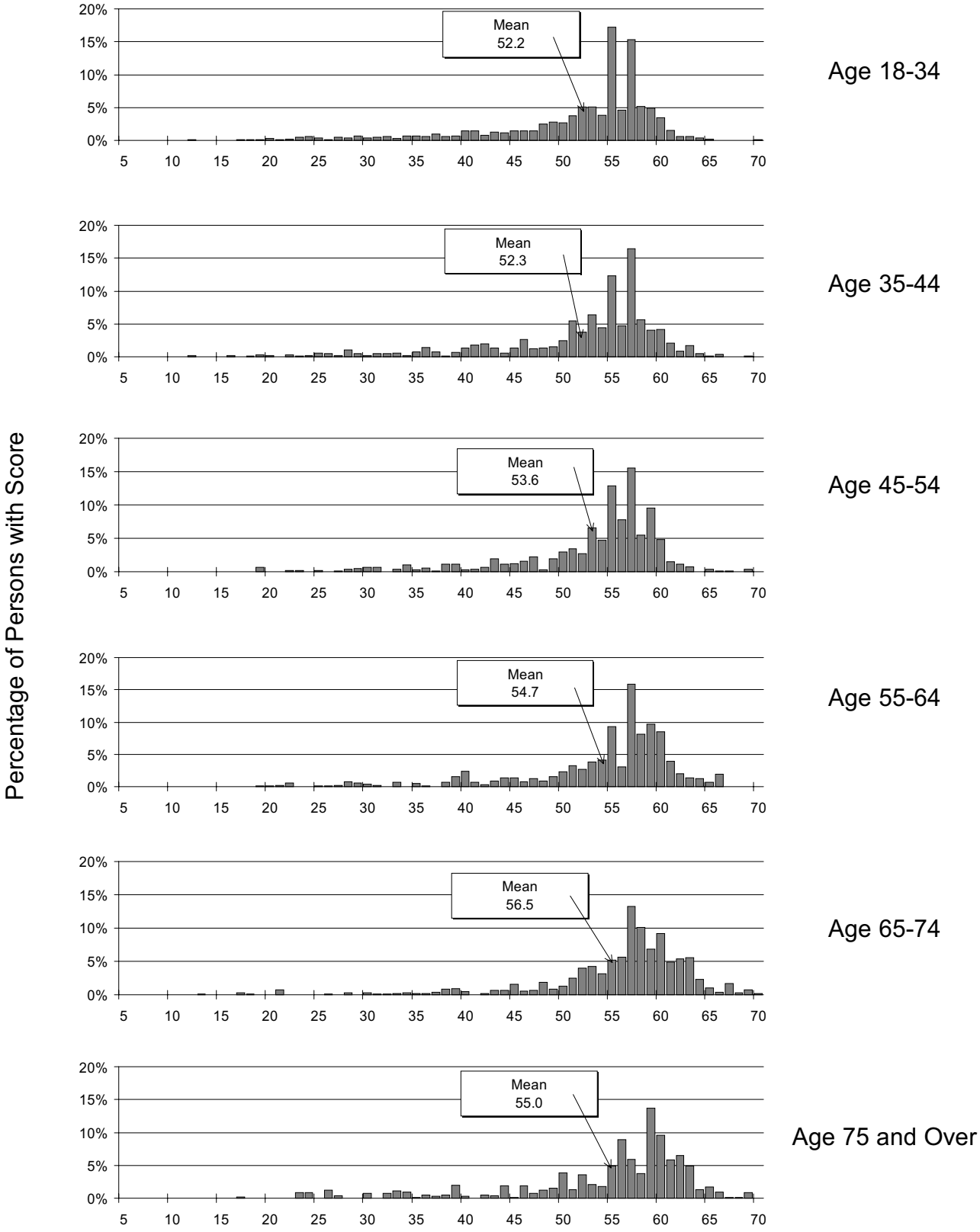
- **Although Utahns' scores on the Physical Health Composite Scale were similar to U.S. scores overall, there was a trend such that younger persons scored lower and older persons scored higher than their U.S. counterparts. Will Utah's health status advantages erode in the future as our younger generations age?**

- **Utahns' scores on the Mental Health Composite Scale are significantly higher than the U.S. scores for every age group.**

Medical Outcomes Study SF12 Physical Health Composite Scale Scores by Age Group: Utahns Age 18 or Over, 1996



Medical Outcomes Study SF12 Mental Health Composite Scale Scores by Age Group: Utahns Age 18 or Over, 1996

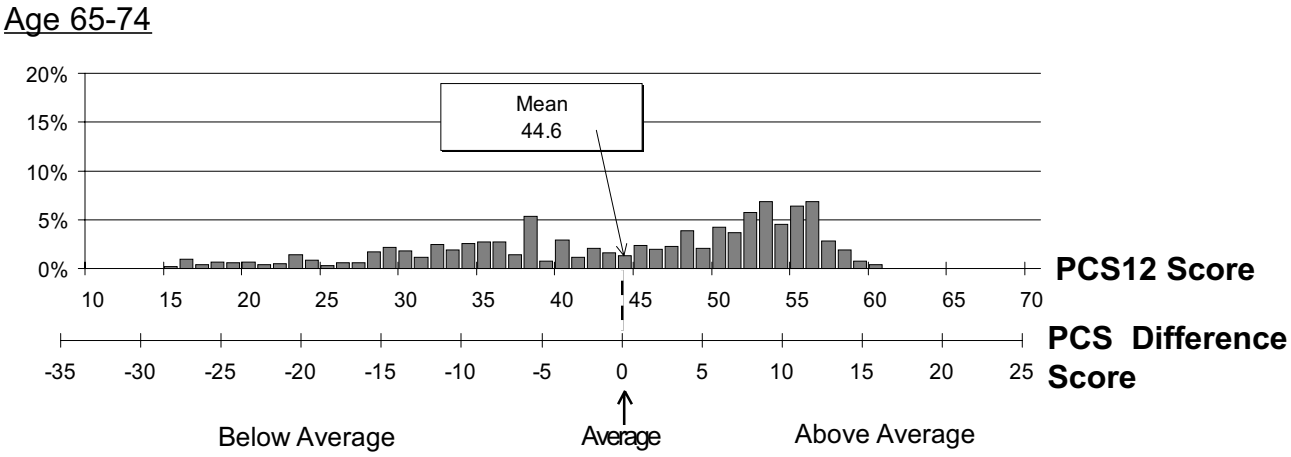


Assigning Meaning to the PCS12 and MCS12 Summary Scores

Computing Difference Scores

- Difference scores can be used to help interpret the meaning of the scale values. The difference score is the difference between a person's score and the mean (average) score for his or her age group.

- A positive score means the person is healthier than average. A negative score means a person is less healthy than average.



Establishing Cut-off Points for Exceptionally Good and Poor Health

In addition to knowing whether a person's score is above or below average, it is also helpful to know whether the score is significantly above or below average. If a person's physical health difference score is negative but very close to zero, they should probably be considered in "average" health. However, if a person's physical health difference score is hovering around minus 20, they should probably be considered "below average," or in poor health. This section explores some methods for deriving cut-off points that define where average ends and below average (and above average) begins.

1. Statistical Methods for Establishing Cut-off Points.

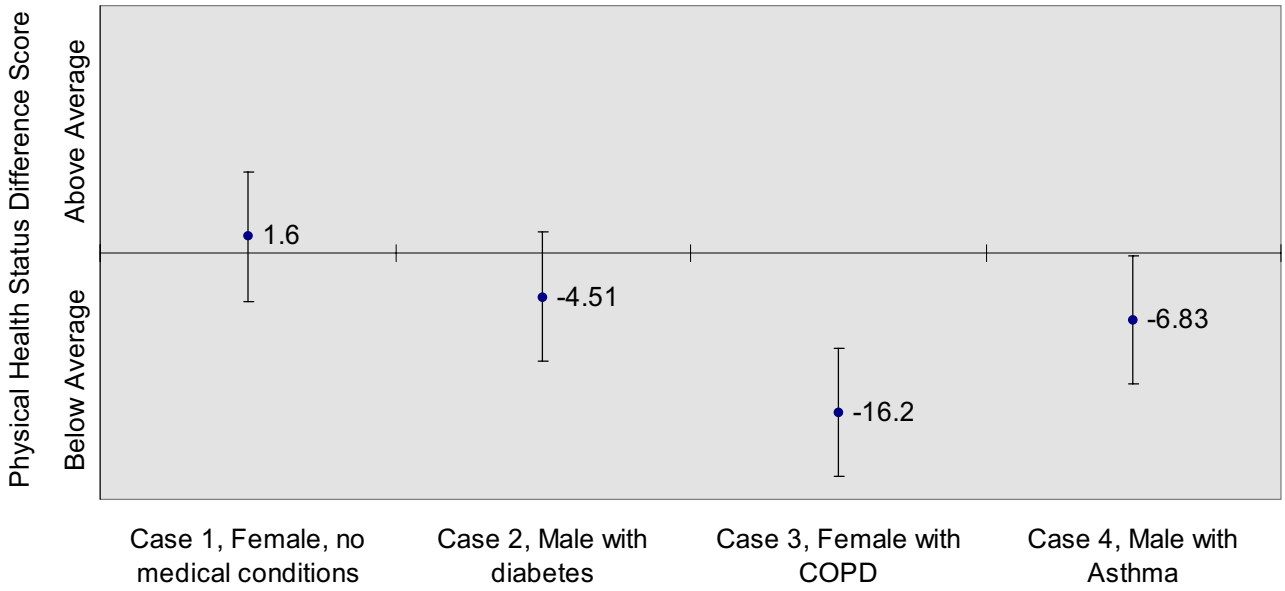
Statistical methods for establishing cut-off points all rely on measures of variability (such as standard deviation and standard error). These measures use confidence intervals to define whether the individual or the group has a score that could be construed as "the same as" average -- if the confidence interval includes the zero point (the average score) then the score is the same as average, if the confidence interval does not include the zero point, then the score is different from (above or below) average.

Cut-off Points for Individual Scores. The cut-off point for an individual's score is based on a property of the SF-12 scale, called the standard error of measurement. All other things being equal, a person's score is expected to have some normal amount of variation that should not be interpreted as a change in the individual's health status. The 95% confidence interval is 1.96 times

the standard error of measurement. The 95% confidence interval was calculated at ±6.53 points for the Physical Health Summary Measure (and also for the Physical Health Status Difference Score), and ±6.11 points for the Mental Health Summary Measure (and Mental Health Status Difference Score).

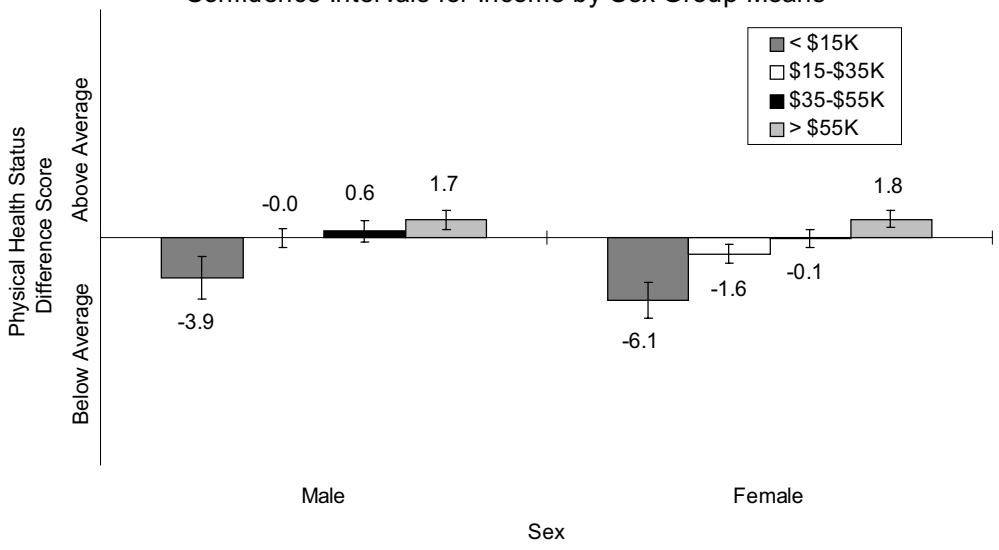
If we apply this approach to an individual's Physical Health Status Difference Score, then the 95% confidence interval for a score that is below -6.53 will not include zero, and will be considered "below average" by this criterion. To demonstrate this process, we can consider the following cases. Case 1 is a female with no chronic medical conditions with a Physical Health Status Difference Score of +1.6, Case 2 is a male diabetic with a score of -4.51, Case 3 is a female with chronic obstructive pulmonary disease with a score of -16.2, and Case 4 is a male with asthma with a score of -6.83. These scores are plotted on Example1 (below) along with the confidence intervals, all of which are ±6.53. Cases 1 and 2 are not significantly different from the average score. Cases 3 and 4, however, are significantly below average.

Example 1. Physical Health Status Difference Scores and Confidence Intervals for Four Hypothetical Individuals



Cut-off Points for Group Means. The mean (or average) score has a measure of deviation, the standard error, that is based on the amount of dispersion or spread of the group's scores around the mean score and the number of persons in that group. Every mean score has it's own standard error. In Example 2 (below) group means and standard errors have been plotted for males and females by income category. Males living in households with less than $15,000 annual income scored significantly below average (the confidence interval does not include zero), males in the middle two income categories had scores in the average range, and males in households with more than $55,000 annual income had physical health status difference scores that were significantly above average.

Example 2. Physical Health Status Difference Scores and Confidence Intervals for Income by Sex Group Means



2. Criterion Methods for Establishing Cut-off Points.

Criterion-based methods for assigning meaning to various physical and mental health summary scores are based on the relationship between those scores and other "criterion" variables measured in the same population. The criteria used here were selected on the basis of their similarity to those used by The Health Institute. Ware et al. (1994) considered these criteria to be both socially and clinically important. They were measured independently of the PCS and MCS scales.

In the following tables, persons in the Utah survey sample were grouped according to their PCS12 and MCS12 Difference Scores, with the healthiest-scoring group assigned to category number 1, and the least healthy-scoring group assigned to category number 8. The average PCS Difference Scores and summary scale scores for these eight groups are reported in the following table.

Comparison of Eight Physical Health Status Groups: PCS12 Difference and Summary Scores

Level	PCS12 Difference Score Range	Weighted Sample Size*	PCS12 Difference Score Mean	PCS12 Summary Scale Mean
1	10 & over	298	11.9 ± *0.3*	59.2 ± *0.5*
2	5 to less than 10	1458	6.7 ± *0.1*	56.5 ± *0.2*
3	0 to less than 5	2210	3.1 ± *0.1*	54.2 ± *0.1*
4	less than 0 to -5	825	-2.2 ± *0.2*	48.5 ± *0.3*
5	less than -5 to -10	513	-7.5 ± *0.2*	42.4 ± *0.5*
6	less than -10 to -15	297	-12.4 ± *0.3*	36.7 ± *0.6*
7	less than -15 to -20	179	-17.0 ± *0.3*	32.7 ± *0.7*
8	less than -20	308	-25.1 ± *0.6*	24.4 ± *0.7*
		6087		

* The sample used here has been weighted to match the age, sex, Hispanic status and geographic distribution of Utah, and then normalized back to the original sample size. It has not been "inflated" to the population size.

Limitations of usual activities and ability to work for pay are used as criterion variables in the table, below. Limitations in usual activities is more sensitive to health status than is ability to work for pay, as there are many who report that they are limited in their usual activities and do not report that they are unable to work. Level 1 has a small number of cases (n = 50), one of whom reported being limited in his or her usual activities.

Judging from this table, a PCS12 Difference Score of <0 would include almost all persons with limitations in usual activities. A more conservative cut-off point would be -10. Below this point many or most persons have some functional limitation. The choice between an inclusive versus a conservative cut-off point would depend on the user's needs.

Comparison of Eight Physical Health Status Groups:
Limitations of Activities and Ability to Work

Level	PCS12 Difference Score Range	Percentage Limited in Usual Activities*	Percentage Reporting "Unable to Work"**
1	10 & over	3.3% ± *6.5%*	0.0% ± *0.0%*
2	5 to less than 10	0.5% ± *0.6%*	0.3% ± *0.4%*
3	0 to less than 5	0.7% ± *0.9%*	0.2% ± *0.3%*
4	less than 0 to -5	13.4% ± *8.1%*	0.8% ± *0.7%*
5	less than -5 to -10	20.2% ± *12.3%*	1.3% ± *1.0%*
6	less than -10 to -15	46.4% ± *20.2%*	7.0% ± *6.1%*
7	less than -15 to -20	45.6% ± *24.5%*	7.6% ± *5.6%*
8	less than -20	73.3% ± *18.2%*	29.8% ± *9.4%*

* Limited in any way in performing usual activities because of an impairment or health problem. Sample n's for this column are approximately one-sixth of the sizes reported in the weighted sample size, above.
** Not employed and unable to work during most of the previous year, adults age 18 to 64.

In the next table, the eight health status groups have been compared on the likelihood of having a chronic medical condition, and medical and hospital visits. Examination of the table reveals that those in the least healthy-scoring groups were significantly more likely to have one or more chronic medical conditions, more doctor visits in the last year, and an overnight hospital stay. The most commonly reported medical condition among the Level 1 group was high blood pressure. The relatively high number of medical visits among the Level 1 members can be accounted for by two survey respondents who reported they had sought mental health services in the last year, and had 50 or more medical visits (possibly mental health visits) during the last 12 months.

The likelihood of a chronic medical condition is 50% or higher for persons with a PCS12 Difference Score below -5. The likelihood of having been hospitalized in the last year increases only for those with a PCS12 Difference Score below -15 or -20. Again, it appears that a cut-off point somewhere near level 5 or 6 is suggested by this criterion-based method.

Comparison of Eight Physical Health Status Groups: Medical Conditions, Medical Visits and Hospitalization

Level	PCS12 Difference Score Range	Percentage With One or More Medical Conditions*	Mean Number of Medical Visits, Last 12 Mos.**	Percentage With One or More Hospital Visits, Last 12 Mos.**
1	10 & over	51.4% ± *8.0%*	4.6 ± *3.7*	1.2% ± *1.3%*
2	5 to less than 10	33.6% ± *3.9%*	2.4 ± *0.5*	2.5% ± *2.5%*
3	0 to less than 5	25.8% ± *2.9%*	2.3 ± *0.6*	1.1% ± *1.0%*
4	less than 0 to -5	46.1% ± *5.5%*	4.0 ± *1.5*	3.9% ± *6.5%*
5	less than -5 to -10	50.8% ± *6.9%*	4.1 ± *3.3*	13.7% ± *18.4%*
6	less than -10 to -15	60.9% ± *8.6%*	5.5 ± *3.4*	8.7% ± *11.0%*
7	less than -15 to -20	64.0% ± *11.5%*	9.4 ± *5.9*	16.3% ± *15.5%*
8	less than -20	72.6% ± *7.7%*	10.2 ± *2.8*	11.7% ± *9.1%*

* Medical conditions include asthma, diabetes, chronic obstructive pulmonary disease, arthritis, stroke, heart disease, high blood pressure, hearing impairment, uncorrectable vision, and speech impairment.
** Sample n's for these columns are approximately one-sixth of the sizes reported in the weighted sample size, above.

The following tables examine the same or similar measures for eight levels of the MCS12 Difference Score, as displayed in the following table.

Comparison of Eight Mental Health Status Groups: MCS12 Difference and Summary Scores

Level	MCS12 Difference Score Range	Weighted Sample Size*	MCS12 Difference Score Mean	MCS12 Summary Scale Mean
1	10 & over	130	11.9 ± *0.4*	65.2 ± *0.6*
2	5 to less than 10	1562	6.6 ± *0.1*	59.6 ± *0.1*
3	0 to less than 5	2414	2.8 ± *0.1*	56.0 ± *0.1*
4	less than 0 to -5	844	-2.2 ± *0.2*	51.2 ± *0.2*
5	less than -5 to -10	390	-7.3 ± *0.3*	45.7 ± *0.3*
6	less than -10 to -15	293	-12.1 ± *0.3*	41.0 ± *0.4*
7	less than -15 to -20	179	-17.2 ± *0.3*	35.8 ± *0.5*
8	less than -20	277	-26.5 ± *0.8*	26.5 ± *0.7*
		6087		

* The sample used here has been weighted to match the age, sex, Hispanic status and geographic distribution of Utah, and then normalized back to the original sample size. It has not been "inflated" to the population size.

The next table examines limitations of activities and ability to work. While there was no predicted covariation between functional limitations and mental health status, we have presented this table for comparison with the physical health status table on page 17. There was a relatively high proportion (34.8%) of persons in Mental Health Status Level 1 who were limited in their usual activities. This number is based on a very small sample size (n=23) and has a correspondingly large confidence interval. The numbers in the rest of this table indicate that there is a relationship between mental health status and limitations in activities, and that it may be reasonable to use the limitations of activities measure as a criterion variable in setting a cut-point for poor mental health. Although this relationship is not as strong as that found between physical health status and limitations in activities, there is still an association between mental health status and limitations in activities, even after controlling for the effects of age and physical health status. The rate of those with limitations in activities increased sharply at level 6 (an MCS12 Difference Score below -10). There is also a weak association between mental health status and ability to work for pay among adults age 18 to 64.

The next table presents clear evidence of an association between mental health status and the

Comparison of Eight Mental Health Status Groups:
Limitations of Activities and Ability to Work

Level	MCS12 Difference Score Range	Percentage Limited in Usual Activities*	Percentage Reporting "Unable to Work"**
1	10 & over	34.8% ± *27.3%*	6.7% ± *5.4%*
2	5 to less than 10	9.3% ± *5.0%*	1.4% ± *1.1%*
3	0 to less than 5	6.6% ± *3.6%*	0.6% ± *0.4%*
4	less than 0 to -5	10.3% ± *7.0%*	1.6% ± *1.4%*
5	less than -5 to -10	12.1% ± *8.9%*	4.4% ± *5.9%*
6	less than -10 to -15	30.5% ± *17.4%*	8.9% ± *5.2%*
7	less than -15 to -20	23.7% ± *22.6%*	7.2% ± *4.2%*
8	less than -20	37.2% ± *20.6%*	8.8% ± *4.4%*

* Limited in any way in performing usual activities because of an impairment or health problem. Sample n's for this column are approximately one-sixth of the sizes reported in the weighted sample size, above.
** Not employed and unable to work during most of the previous year, adults age 18 to 64.

percentage of persons seeking mental health services in the last 12 months. Mental health help-seeking behavior is a variable that has strong conceptual face-validity for use as a criterion variable for setting a cut-point for defining poor mental health. The data in the following table suggest three levels of mental health status: Good or excellent mental health status is indicated by small percentages of persons seeking help in levels 1 through 4. In levels 5, 6 and 7, about a fifth of persons sought mental health services. And in level 8, at MCS12 Difference Scores less than -20, about two-fifths of persons sought mental health services in the last 12 months. Conventional wisdom suggests that there are still social barriers to seeking mental health services, and that those who actually seek help are representative of larger numbers who would benefit from help but do not seek it. If this is true, perhaps a cut-off point at level 5, at an MCS12 Difference Score below -5, is indicated by these data. There was a weak association between MCS12 Difference Scores and the number of medical visits, and virtually no relationship was found for the likelihood of a hospital visit in the last 12 months.

Reviewing the evidence from the eight-level groupings of both the PCS12 and MCS12

Comparison of Eight Mental Health Status Groups: Mental Health Help-Seeking Behavior, Medical Visits and Hospitalization

Level	MCS12 Difference Score Range	Percentage Seeking Mental Health Services in Last 12 Months	Mean Number of Medical Visits, Last 12 Mos.*	Percentage With One or More Hospital Visits, Last 12 Mos.*
1	10 & over	8.5% ± *8.1%*	6.5 ± *4.9*	16.5% ± *22.1%*
2	5 to less than 10	3.1% ± *1.3%*	3.9 ± *1.5*	6.2% ± *6.0%*
3	0 to less than 5	4.4% ± *1.2%*	2.8 ± *0.6*	1.7% ± *1.5%*
4	less than 0 to -5	9.0% ± *2.9%*	3.5 ± *1.4*	3.2% ± *3.4%*
5	less than -5 to -10	18.5% ± *5.8%*	2.9 ± *1.2*	11.2% ± *13.6%*
6	less than -10 to -15	19.1% ± *7.8%*	3.1 ± *1.4*	4.4% ± *4.9%*
7	less than -15 to -20	25.7% ± *9.7%*	7.6 ± *6.7*	18.3% ± *24.4%*
8	less than -20	41.5% ± *8.3%*	7.1 ± *4.0*	1.2% ± *1.9%*

* Sample n's for these columns are approximately one-sixth of the sizes reported in the weighted sample sizes for the MCS difference scores, above.

Difference Scores, it appears that a cut-off point for defining poor health status could be drawn as high as -5 (at level 5). This seems to be the level at which health status begins to effect various aspects of a person's life. It is also very similar to the cut-off points for individual scores, described on pages 14 and 15. With a PCS12 Difference Score below -5, more than half have one or more chronic medical conditions and about 20% have a functional limitation in their usual activities. Similarly, persons with an MCS12 Difference Score below -5 have about a 20% probability of having sought mental health services in the last 12 months. It is only at much lower difference score levels that a person's physical or mental health status appears to have grave impacts on his or her life. At a PCS12 Difference Score level below -20, a person has about a 73% chance of having a functional limitation, and about a 30% chance of being disabled and unable to work. Also at this level, about 73% have one or more chronic medical conditions. At an MCS12 Difference Score level below -20, about 40% have sought mental health services in the last 12 months.

The question of which level to use as a cut-off point for defining poor health status depends partially on one's use of the data, that is, one's reason for asking. If the user is trying to identify a group of persons whose lives are certainly and severely affected by their health status, then the user should use a cut-off point around -20. However, for most purposes, a less restrictive cut-off point is desirable because it identifies persons whose lives are probably being affected by their health, even though it wouldn't always be manifest on one of a limited number of criterion measures. These data suggest that the method described on pages 14 and 15, that is, PCS12 or MCS12 Differences Scores below -6.53 (physical health) or -6.11 (mental health) for use as this less restrictive level.